Multi-wavelength imaging algorithm for optical interferometry

Éric Thiébaut, Ferréol Soulez

▶ **To cite this version:**

Éric Thiébaut, Ferréol Soulez. Multi-wavelength imaging algorithm for optical interferometry. SPIE Astronomical Telescopes and Instrumentation, Jul 2012, Amsterdam, Netherlands. pp.34. hal-00715809

HAL Id: hal-00715809
https://hal.science/hal-00715809

Submitted on 9 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multi-wavelength imaging algorithm for optical interferometry

Éric Thiébaut[a] & Ferréol Soulez[a]

[a]Université de Lyon, Lyon, F-69003, France; Université Lyon 1, Observatoire de Lyon, 9 avenue Charles André, Saint-Genis Laval, F-69230, France; CNRS, UMR 5574, Centre de Recherche Astrophysique de Lyon; École Normale Supérieure de Lyon, Lyon, F-69007, France.

ABSTRACT

Optical interferometers provide multiple wavelength measurements. In order to fully exploit the spectral and spatial resolution of these instruments, new algorithms for image reconstruction have to be developed. Early attempts to deal with multi-chromatic interferometric data have consisted in recovering a gray image of the object or independent monochromatic images in some spectral bandwidths. The main challenge is now to recover the full 3-D (spatio-spectral) brightness distribution of the astronomical target given all the available data. We describe a new approach to implement multi-wavelength image reconstruction in the case where the observed scene is a collection of point-like sources. We show the gain in image quality (both spatially and spectrally) achieved by globally taking into account all the data instead of dealing with independent spectral slices. This is achieved thanks to a regularization which favors spatially sparsity and spectral grouping of the sources. Since the objective function is not differentiable, we had to develop a specialized optimization algorithm which also takes into account the non-negativity of the brightness distribution.

Keywords: optical interferometry; image reconstruction; inverse problem; sparsity.

1. CONTEXT

The objective of stellar interferometric imaging is to recover an approximation the specific brightness distribution $I_\lambda(\boldsymbol{\theta})$ of the observed astronomical object given measurements which are provide sparse samples of the spatial Fourier transform of $I_\lambda(\boldsymbol{\theta})$. The reconstruction of a monochromatic image from optical interferometry data is a challenging task which has been the subject of fruitful research and resulted in various algorithms (*e.g.*, MiRA,[1] BSMEM,[2,3] WISARD,[4] *etc.*). When dealing with multi-spectral data, a first possibility is to process each wavelength independently and reconstruct a monochromatic image for each subset of measurements from a given spectral channel. For instance, this is what have been done by le Bouquin et al.[5] for the multi-spectral images of the Mira star T Lep. Another possibility is to exploit some assumed spectral continuity of $I_\lambda(\theta)$ and process the multi-spectral data globally to reconstruct an approximation of the 3-D distribution $I_\lambda(\theta)$. Although demonstrated in a different context of integral field spectral spectroscopy, this latter approach has proven to be more powerful[6,7] and this paper aims at exploiting the advantages of a global multi-spectral processing of optical interferometric data.

In order to simplify the problem, we restricted our study to the cases where the complex visibilities are observed and where the observed scene is a collection of point-like sources. In some sense, this latter assumption makes our algorithm a successor of the CLEAN algorithm[8,9] and the *building-block method*[10] developed for recovering monochromatic images from radio and optical interferometric data respectively. In addition to processing multi-variate data, we however introduce the explicit minimization of a non-differentiable regularization term so as to favor spatial sparsity of the reconstructed brightness distribution in a way which is known to be more efficient[11,12] than *greedy algorithms* like CLEAN[13] or the *building-block method*.[10]

Our paper is organized as follows: we first summarize the inverse approach for image reconstruction and discuss various possibilities to impose spatial sparsity, we then detail our algorithm for minimizing the objective function; finally we present some results on simulated data and discuss the advantages of our approach.

Further author information: thiebaut@obs.univ-lyon1.fr

2. ALGORITHMS

2.1 Principle of image reconstruction

Following inverse approach described by, *e.g.* Thiébaut,[14] we state image reconstruction as a constrained optimization problem:

$$\boldsymbol{x}^+ = \underset{\boldsymbol{x}\in\mathbb{X}}{\arg\min} \left\{ f_{\mathsf{data}}(\boldsymbol{x}|\boldsymbol{y}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) \right\} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^{\mathrm{Card}(\boldsymbol{x})}$ are the sought image parameters, $\mathrm{Card}(\boldsymbol{x})$ is the number of parameters, $\mathbb{X} \subset \mathbb{R}^{\mathrm{Card}(\boldsymbol{x})}$ is the set of acceptable parameters, $\boldsymbol{y} \in \mathbb{R}^{\mathrm{Card}(\boldsymbol{y})}$ are the measurements, $f_{\mathsf{data}}(\boldsymbol{x}|\boldsymbol{y})$ is a likelihood term which enforces agreement of the model with the data, $f_{\mathsf{prior}}(\boldsymbol{x})$ is a regularization term which introduces *a priori* constraints and $\mu > 0$ is a so-called hyper-parameter used to tune the relative weight of the prior constraints.

Due to the voids in the $u-v$ coverage of the observations, the constraints set by the data alone do not permit to define a unique image. The prior constraints are then required to help choosing a unique solution among all the images that are compatible with the measurements.

Constraining the solution to belong to the feasible set $\mathbb{X}$ is a mean to impose *strict constraints* such as the positivity:

$$\mathbb{X} = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq 0\} \tag{2}$$

where the inequality $\boldsymbol{x} \geq 0$ is taken element-wise.

2.2 Likelihood and direct model

2.2.1 Monochromatic case

In this preliminary study, we purposely omit some of the difficulties specific to optical interferometry: we do not impose any normalization constraints on the sought distribution and we assume that complex visibilities have been measured (*e.g.*, phase-referenced observations). In this context, assuming Gaussian noise distribution, the likelihood term writes:

$$f_{\mathsf{data}}(\boldsymbol{x}|\boldsymbol{y}) = (\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{y})^\top \cdot \mathbf{W} \cdot (\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{y}) \tag{3}$$

where $\mathbf{H} \in \mathbb{R}^{\mathrm{Card}(\boldsymbol{y})\times\mathrm{Card}(\boldsymbol{x})}$ is the linear model matrix and $\mathbf{W} \in \mathbb{R}^{\mathrm{Card}(\boldsymbol{y})\times\mathrm{Card}(\boldsymbol{y})}$ is a statistical weighting matrix, in principle $\mathbf{W}$ is the inverse of the covariance matrix of the measurements: $\mathbf{W} = \mathrm{Cov}(\boldsymbol{y})^{-1}$.

To avoid dealing with complex arithmetic, we simply consider that complex values are just pairs of real. Hence the total number of observed spatial frequencies is $\approx m/2$. For monochromatic interferometric data, the model matrix then writes:

$$H_{m,n} = \begin{cases} +\cos(\boldsymbol{\theta}_n^\top \cdot \boldsymbol{b}_m/\lambda) & \text{if } y_m \text{ is the real part of a measured complex visibility} \\ -\sin(\boldsymbol{\theta}_n^\top \cdot \boldsymbol{b}_m/\lambda) & \text{if } y_m \text{ is the imaginary part of a measured complex visibility} \end{cases} \tag{4}$$

with $\boldsymbol{b}_m$ the observing baseline for the mth datum, λ the wavelength and $\boldsymbol{\theta}_n$ the angular direction of nth picture element (pixel):

$$x_n \approx I_\lambda(\boldsymbol{\theta}_n)\,.$$

In Eq. (4), $\boldsymbol{\theta}_n^\top \cdot \boldsymbol{b}_m$ is the usual scalar product between the baseline vector $\boldsymbol{b}_m$ and the view direction $\boldsymbol{\theta}_n$.

At least because of the strict constraints imposed by the feasible set $\mathbb{X}$, that is whatever the type of regularization, solving the image reconstruction problem in Eq. (1) must be done by an iterative optimization algorithm. Thus, owing to the size of the problem, it is strongly recommended to use a fast approximation of $\mathbf{H}$. The *nonuniform fast Fourier transform*[15] (NU-FFT) is a good choice for that purpose. In this case, the operator $\mathbf{H}$ has the following structure:

$$\mathbf{H} = \mathbf{R}\cdot\mathbf{F}\cdot\mathbf{S} \tag{5}$$

where $\mathbf{F}$ is the discrete Fourier transform (DFT), computed by a fast Fourier transform (FFT) algorithm, $\mathbf{R}$ interpolates the discrete spatial frequencies resulting form the DFT at the frequencies effectively observed and $\mathbf{S}$ is an apodizing operator (usually $\mathbf{S}$ is diagonal) which pre-compensates for the convolution by the interpolation kernel used in $\mathbf{R}$.

2.2.2 Multi-Spectral case

Although in principle any scheme is allowed for sampling a multi-variate distribution such as $I_\lambda(\theta)$, it is usually more practical to sample the sought distribution separately along its spatial and spectral dimensions. Hence, for multi-spectral image reconstruction we consider that the image parameters are:

$$x_{n,\ell} \approx I_{\lambda_\ell}(\theta_n) \tag{6}$$

for $\lambda_\ell \in \mathbb{W}$ the list of sampled wavelengths and $\boldsymbol{\theta}_n \in \mathbb{A}$ the list of angular directions corresponding to the pixels. For the sake of simplicity, we use the same wavelengths in $\mathbb{W}$ as the ones of the observed spectral channels and we use the notation $y_{m,\ell}$ for the multi-spectral data obtained with mth baseline in ℓth spectral. This notation is intended to clarify the equations and does not impose or assume that all baselines have been observed in all spectral channel. Note that since the both real and imaginary parts are measured, each baseline appears at least twice with this convention.

In our framework, the model of the observed complex visibilities writes:

$$y_{m,\ell} \approx \sum_n H_{m,n,\ell}\, x_{n,\ell}\,. \tag{7}$$

The model is therefore separable along the spectral dimension. With the generalization of multi-processor computers or multi-core processors, this property of the operator $\mathbf{H}$ may be exploited to parallelize the code to apply $\mathbf{H}$ (or its adjoint $\mathbf{H}^\top$) in a very straightforward way. Using a true matrix representation, $\mathbf{H}$ would have a block diagonal structure. Formally, the coefficients of the multi-spectral version of $\mathbf{H}$ are given by:

$$H_{m,n,\ell} = \begin{cases} +\cos(\boldsymbol{\theta}_n^\top \cdot \boldsymbol{b}_m/\lambda_\ell) & \text{if } y_{m,\ell} \text{ is the real part of a measured complex visibility} \\ -\sin(\boldsymbol{\theta}_n^\top \cdot \boldsymbol{b}_m/\lambda_\ell) & \text{if } y_{m,\ell} \text{ is the imaginary part of a measured complex visibility} \end{cases} \tag{8}$$

In order to speed up computations, the monochromatic operator $H_{\cdot\,\cdot\,\ell}$ can be applied using the NU-FFT algorithm.

2.3 Regularization by assuming spatial sparsity

In this paper, we focus on a particular type of astronomical targets which consist in group of point like sources. These sources may have different spectra. For instance, this is suitable for multiple stars or group of stars like in globular clusters or in the center of our galaxy. For such objects, the most effective means to regularize the problem is to favor spatially sparse distributions, that is images with as few as possible sources on a dark background. In this section, we derive expressions of the regularization term $f_{\mathsf{prior}}(\boldsymbol{x})$ suitable to favor spatially sparse distributions.

2.3.1 Separable sparsity

It is now well known that good means to impose the sparsity of the solution in the parameter space (which means that most model parameters will be equal to zero) while approximating the data, is to use the ℓ_1 norm as the regularization term:[11]

$$f_{\mathsf{sparse}}(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 \stackrel{\text{def}}{=} \sum_{k,\ell} |x_{k,\ell}| = \mathbf{1}^\top \cdot \boldsymbol{x}\,, \tag{9}$$

where the last equality occurs because of the positivity of the parameters. In principle, any $f_{\mathsf{prior}}(\boldsymbol{x}) = \|\boldsymbol{x}\|_p^p$, with $\|\boldsymbol{x}\|_p$ the ℓ_p norm of the sought parameters and $0 \leq p \leq 1$, can be used to get a sparse distribution with $p = 0$ corresponding the most to a true sparsity prior. However $p = 1$ is the smallest p which yields a convex criterion and therefore guarantees a unique solution. In practice, the ℓ_1 norm is very effective to obtain a sparse solution.[11]

The regularization term $f_{\mathsf{sparse}}(\boldsymbol{x})$ in Eq. (9) is completely separable. Hence, in our framework where the model is spectrally separable, the criterion defined in Eq. (1) is also separable along the spectral dimension. Thus the image reconstruction can be solved independently for each spectral channel. We therefore do not expect as good

results in this case compared to reconstructions performed while imposing some spectral continuity. We shall however use this fully separable regularization in our tests to demonstrate the benefit of imposing non-separable priors along the spectral dimension.

Note that, if the wavelength samples used in the discrete model $\boldsymbol{x}$ of $I_\lambda(\theta)$ do not coincide with the effective wavelengths of the observed spectral channel, spectral interpolation of the model has to be used to match the observed wavelengths. In this case, the model implicitly set some spectral correlation and the 3-D image reconstruction has to be performed globally even if the regularization does not impose any kind of spectral continuity.

2.3.2 Non-separable spatial-only sparsity

The following regularization favors spatial sparsity but with sources at the same location whatever the wavelength:[6]

$$f_{\mathsf{group}}(\boldsymbol{x}) = \sum\nolimits_n \left(\sum\nolimits_\ell x_{n,\ell}^2 \right)^{1/2} \tag{10}$$

where n is the spatial index (pixel) and ℓ is the spectral channel. The fact that such a regularization favors spatial sparsity and spectral grouping is a consequence of the triangular inequality inequality.[12]

2.3.3 Means to impose spectral continuity

Some authors[6,7,16] have shown the efficiency of exploiting the spectral continuity of the sources by using a spectral regularization like:

$$f_{\mathsf{spectral}}(\boldsymbol{x}) = \sum_n \mu_n \sum_\ell (x_{n,\ell+1} - x_{n,\ell})^2 \tag{11}$$

with $\mu_n > 0$ suitable normalization weights.

A rather extreme means to impose spectral continuity is to assume that the spectral energy distribution is everywhere flat. Using the regularization in Eq. (11), this correspond to the limit $\mu_n \to \infty, \forall n$, and yields:

$$x_{n,\ell} = g_n \,, \quad \forall \ell \tag{12}$$

where g is a *gray image* of the object. To speed up the reconstruction, only the gray image has to be reconstructed, using the model:

$$y_{m,\ell} \approx \sum_n H_{m,n,\ell}\, g_n \,, \tag{13}$$

and, to impose the spatial sparsity, the following regularization:

$$f_{\mathsf{prior}}(g) = \|g\|_1 = \sum\nolimits_n |g_n| = \mathbf{1}^\top \cdot \boldsymbol{g} \,, \tag{14}$$

since $\boldsymbol{g} \geq \mathbf{0}$.

2.4 Optimization algorithm

MiRA algorithm[1] was designed for minimizing a smooth cost function. For that purpose, limited memory quasi-Newton methods such as VMLM-B[17] are relatively efficient and easy to use as they only require the cost function and its gradient to be computed. When using non smooth regularizations to impose spatial sparsity as the ones in Eq. (9) and Eq. (10), optimization algorithms based on Newton method (that is, on a quadratic approximation of the cost function) are inefficient and we had to implement a completely different optimization strategy to solve the problem in Eq. (1) with non-differentiable cost functions. In our algorithm, we introduce variables splitting[18] to treat the two terms of our cost function as independently as possible and exploit the alternating direction method of multipliers[19] (ADMM) to solve the resulting constrained problem.

Introducing auxiliary variables $\boldsymbol{z}$, minimizing the two-term cost function in Eq. (1) can be recast in the following equivalent constrained problem:

$$\min_{\boldsymbol{x} \in \mathbb{X}, \boldsymbol{z}} \{ f_{\mathsf{data}}(\boldsymbol{z}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) \} \quad \text{s.t.} \quad \boldsymbol{x} = \boldsymbol{z} \,. \tag{15}$$

Imposing that $\boldsymbol{x} \in \mathbb{X}$ (rather than $\boldsymbol{z} \in \mathbb{X}$) is not arbitrary and our motivation for that choice will be clarified in what follows.

The augmented Lagrangian[20] of this problem writes:

$$\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) = f_{\mathsf{data}}(\boldsymbol{z}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) + \boldsymbol{u}^\top \cdot (\boldsymbol{x} - \boldsymbol{z}) + \frac{\rho}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 \,, \tag{16}$$

with $\boldsymbol{u}$ the Lagrange multipliers associated with the constraints that $\boldsymbol{x} = \boldsymbol{z}$, $\rho > 0$ the quadratic weight of the constraints, and $\|\boldsymbol{v}\|_2$ the Euclidean (ℓ_2) norm of $\boldsymbol{v}$.

The alternating direction method of multipliers[19] (ADMM) consists in alternatively minimizing the augmented Lagrangian for $\boldsymbol{x}$ given $\boldsymbol{z}$ and $\boldsymbol{u}$, then for $\boldsymbol{z}$ given $\boldsymbol{x}$ and $\boldsymbol{u}$, and finally updating the multipliers $\boldsymbol{u}$. This scheme, adapted to our specific problem in Eq. (15), is detailed in the following algorithm with the convention that $\boldsymbol{v}^{(t)}$ is the value of $\boldsymbol{v}$ at iteration number t.

Algorithm 1. *Resolution of Eq. (15) by alternating direction method of multipliers.* Given $\rho > 0$, choose initial variables $\boldsymbol{z}^{(0)}$ and Lagrange multipliers $\boldsymbol{u}^{(0)}$. Then repeat, for $t = 0, 1, 2, \ldots$ until convergence, the following steps:

a/ update variables $\boldsymbol{x}$:

$$\begin{aligned} \boldsymbol{x}^{(t+1)} &= \underset{\boldsymbol{x}\in\mathbb{X}}{\arg\min}\, \mathcal{L}_\rho\left(\boldsymbol{x}, \boldsymbol{z}^{(t)}, \boldsymbol{u}^{(t)}\right) \\ &= \underset{\boldsymbol{x}\in\mathbb{X}}{\arg\min} \left\{ f_{\mathsf{prior}}(\boldsymbol{x}) + \frac{\rho}{2\,\mu} \left\| \boldsymbol{x} - \widetilde{\boldsymbol{x}}^{(t)} \right\|_2^2 \right\} \qquad \text{with:} \quad \widetilde{\boldsymbol{x}}^{(t)} = \boldsymbol{z}^{(t)} - \boldsymbol{u}^{(t)}/\rho \,; \end{aligned} \tag{17}$$

b/ update variables $\boldsymbol{z}$:

$$\begin{aligned} \boldsymbol{z}^{(t+1)} &= \underset{\boldsymbol{z}}{\arg\min}\, \mathcal{L}_\rho\left(\boldsymbol{x}^{(t+1)}, \boldsymbol{z}, \boldsymbol{u}^{(t)}\right) \\ &= \underset{\boldsymbol{z}}{\arg\min} \left\{ f_{\mathsf{data}}(\boldsymbol{z}) + \frac{\rho}{2} \left\| \boldsymbol{z} - \widetilde{\boldsymbol{z}}^{(t)} \right\|_2^2 \right\} \qquad \text{with:} \quad \widetilde{\boldsymbol{z}}^{(t)} = \boldsymbol{x}^{(t+1)} + \boldsymbol{u}^{(t)}/\rho \,; \end{aligned} \tag{18}$$

c/ update multipliers $\boldsymbol{u}$:

$$\boldsymbol{u}^{(t+1)} = \boldsymbol{u}^{(t)} + \rho \left(\boldsymbol{x}^{(t+1)} - \boldsymbol{z}^{(t+1)} \right) \tag{19}$$

In Appendix A, we demonstrate how the updating of the variables $\boldsymbol{x}$ (step a of the algorithm) can be done efficiently and taking into account the constraint that the parameters are positive. This is our motivation for imposing $\boldsymbol{x} \in \mathbb{X}$ on the variables $\boldsymbol{x}$ and not on the variables $\boldsymbol{z}$ to which the likelihood term $f_{\mathsf{data}}(\boldsymbol{z})$ is applied. Another possibility could have been to introduce additional auxiliary variables for the purpose of accounting for the feasible set. The formulae to update the variables $\boldsymbol{x} \in \mathbb{X}$ for $f_{\mathsf{sparse}}(\boldsymbol{x})$ and $f_{\mathsf{group}}(\boldsymbol{x})$ are respectively given by Eq. (28) and Eq. (34).

The updating of variables $\boldsymbol{z}$ is equivalent to solving a linear problem of which the analytical solution is:

$$\boldsymbol{z}^{(t+1)} = \left(\mathbf{H}^\top \cdot \mathbf{W} \cdot \mathbf{H} + \rho\, \mathbf{I}\right)^{-1} \cdot \left(\mathbf{H}^\top \cdot \mathbf{W} \cdot \boldsymbol{y} + \rho\, \boldsymbol{x}^{(t)} + \boldsymbol{u}^{(t)}\right) \tag{20}$$

with $\mathbf{I}$ the identity matrix (of suitable size). Since the augmented term is diagonal, the spectral separability is preserved and this problem can be solved independently for each spectral channel:

$$\boldsymbol{z}_\ell^{(t+1)} = \mathbf{A}_\ell^{-1} \cdot \boldsymbol{b}_\ell^{(t)} \tag{21}$$

with:

$$\mathbf{A}_\ell = \mathbf{H}_\ell^\top \cdot \mathbf{W}_\ell \cdot \mathbf{H}_\ell + \rho\,\mathbf{I} \tag{22}$$

$$\boldsymbol{b}_\ell^{(t)} = \mathbf{H}_\ell^\top \cdot \mathbf{W}_\ell \cdot \boldsymbol{y}_\ell + \rho\,\boldsymbol{x}_\ell^{(t)} + \boldsymbol{u}_\ell^{(t)} \tag{23}$$

where the index ℓ denotes the sub-vector or the sub-operator restricted to the coefficients corresponding to the ℓth spectral channel. In practice and owing to the number of parameters, we (approximately) solve these problems by means of the conjugate gradients algorithm[21] with left-hand side matrix $\mathbf{A}_\ell$ and right-hand side vector $\boldsymbol{b}_\ell^{(t)}$.

One of the important settings of the ADMM method is to choose properly ρ: if it is too small, the constraints $\boldsymbol{x} = \boldsymbol{z}$ will converge too slowly; while, if it is too small, the cost function will decrease too slowly. The most effective schemes consist in tuning ρ at every iteration of ADMM.[19]

3. RESULTS

To check the proposed algorithm, we simulated a cluster of 50 stars with random positions and luminosities and with spectra randomly taken from the library provided by Jacoby et al.[22] We used 100 spectral channels from $\lambda = 493\,\text{nm}$ to $\lambda = 507\,\text{nm}$ by steps of $\Delta\lambda = 0.14\,\text{nm}$. The field of view is 128×128 pixels of size 0.5 milliarcseconds. To simulate the observations, we took 100 random interferometric baselines with a maximum baseline of 180 m. We added Gaussian white noise to the complex visibilities with a level such that the maximum signal-to-noise ratio (SNR) is equal to 100.

For the image reconstructions, we considered three different cases: the reconstruction of a multi-spectral distribution with the regularization $f_{\mathsf{sparse}}(\boldsymbol{x})$ in Eq. (9), or the regularization $f_{\mathsf{group}}(\boldsymbol{x})$ in Eq. (10), and the reconstruction of a gray object $\boldsymbol{g}$ with the regularization $f_{\mathsf{sparse}}(\boldsymbol{g})$. In order to set the relative weight of the priors, we choose the value of the hyper-parameter μ which yields an image which has the least mean quadratic differences with the true distribution. Once chosen the values of μ and ρ, the reconstruction of a $128 \times 128 \times 100$ distribution from $\sim 2 \times 10^4$ measurements takes a few minutes on a GNU/Linux workstation with a quad-core processor at 3 GHz and using a multi-treaded version of FFTW[23] to compute the discrete Fourier transforms.

Figure 1 shows the integrated flux, *i.e.* $\sum_\ell x_{n,\ell}$, for the true distribution and for the reconstructed ones. In any cases, the spatial sparsity priors effectively yield a solution with point-like structures. However, when there is no trans-spectral constraints, only a few sources are correctly found and there are many more spurious sources. When using $f_{\mathsf{group}}(\boldsymbol{x})$ or assuming a gray object the estimated integrated luminosity is much more consistent with that of the true object: all existing sources are found and the spurious sources are not only less numerous but also much fainter than the true ones.

Figure 2 shows the spectra of the two stars encircled by boxes in Fig. 1, clearly the spectra recovered with $f_{\mathsf{group}}(\boldsymbol{x})$ (blue curves) are of much higher quality than the spectra estimated when treating the spectral channels independently (red curves). Compared to the true spectra (green curves) there is however a small but significant bias in the spectra obtained with $f_{\mathsf{group}}(\boldsymbol{x})$. This is not unexpected as the mixed norm implemented by $f_{\mathsf{group}}(\boldsymbol{x})$ results in a soft compression as shown by Eq. (32). There are different possibilities to avoid this bias. The simplest one consists in using the distribution reconstructed with $f_{\mathsf{group}}(\boldsymbol{x})$ to select the significant sources and then restrict the reconstruction to be non-zero only for these sources (there is no needs for other regularization) REFS. According to the good estimation of the integrated luminosity obtained by imposing a gray object, this may be used to select the location of the sources.

4. DISCUSSION AND PERSPECTIVES

We have shown the importance of using trans-spectral constraints to improve the quality of the restoration of the multi-spectral brightness distribution $I_\lambda(\boldsymbol{\theta})$ of an astronomical target from optical interferometric data. These results confirm what has been observed for other type of data (like integral field spectroscopy).

For the moment, our demonstration is restricted to specific objects which are spatially sparse (*e.g.* point-like sources) and must be generalized to other type of spatial distributions. Being implemented by non-differentiable cost functions, spatial sparsity requires specific optimization algorithms. We have shown that variable splitting

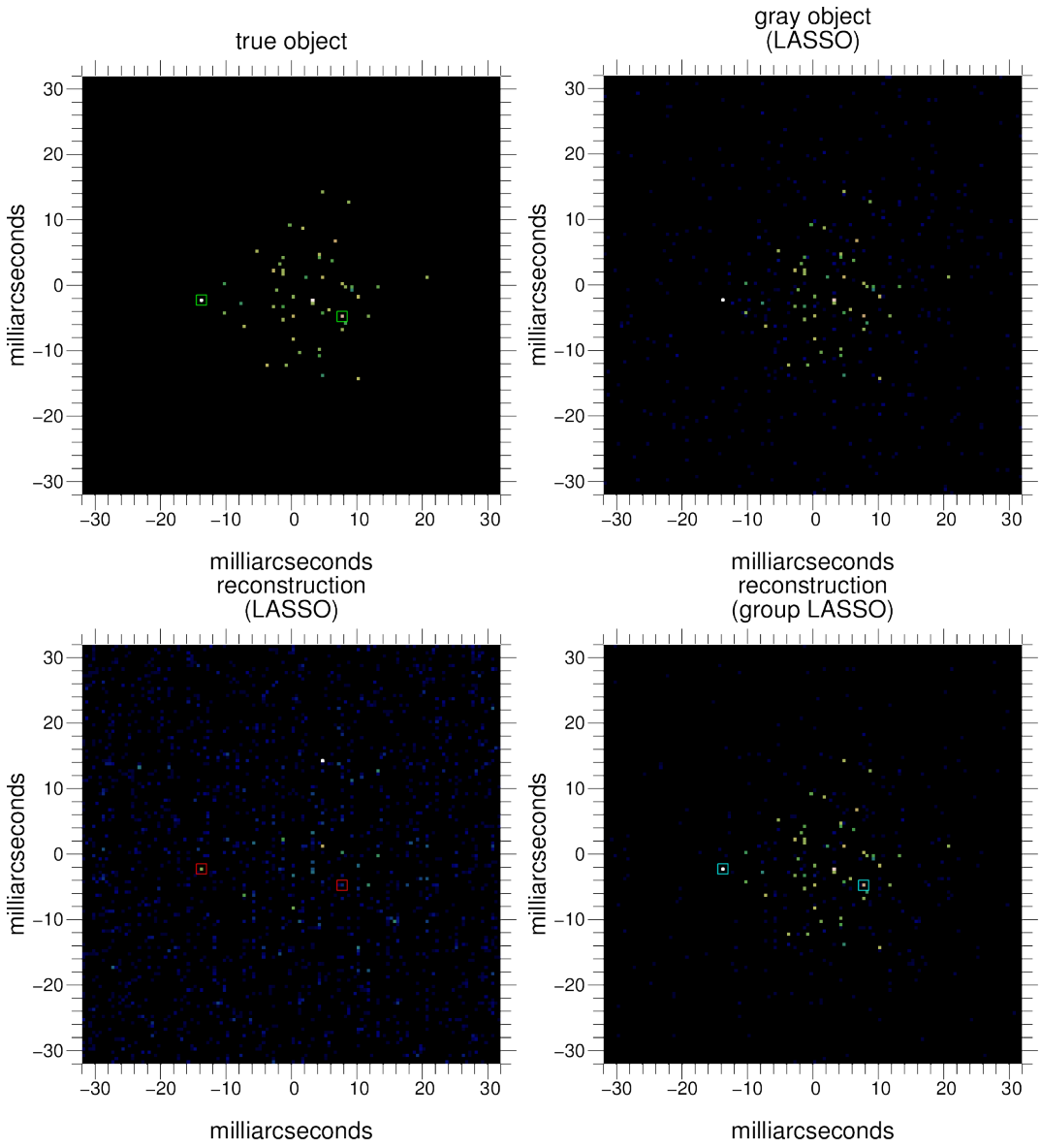

Figure 1. Integrated flux for the star cluster. From top to bottom, left to right: true object; reconstruction with spatial sparsity and assuming a gray object; reconstruction with fully separable sparsity prior; reconstruction with semi-separable sparsity prior. The spectra of the sources encircled by the boxes are shown in Fig. 2.

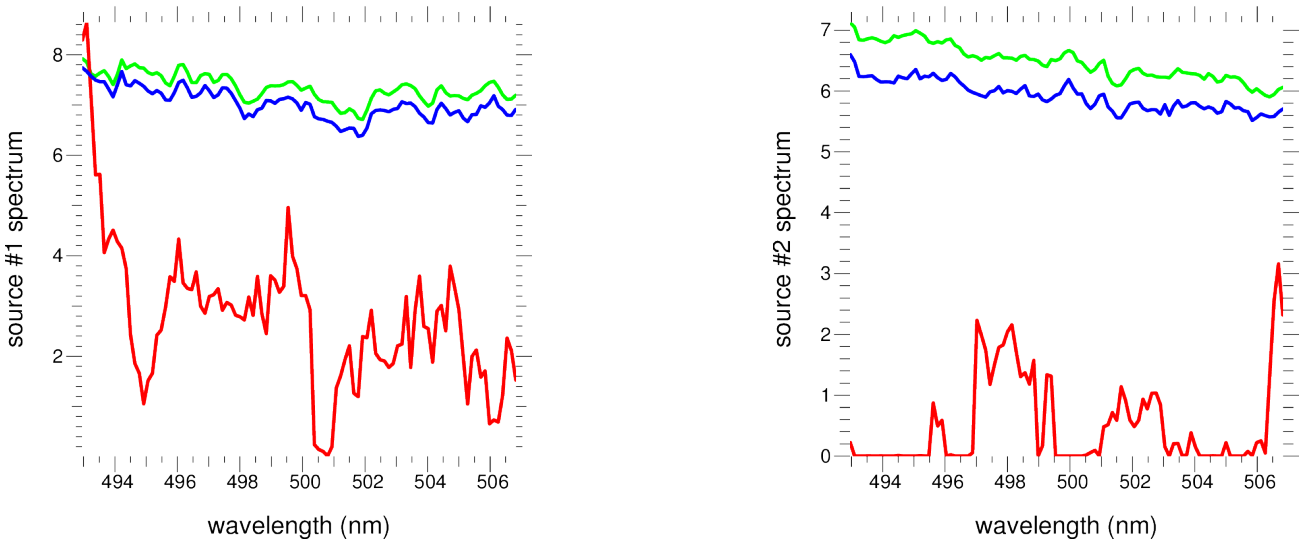

Figure 2. Spectra of two selected sources. Each panel show the spectra of one of the sources encircled by the boxes in Fig. 1. In *green*: the true spectra; in *red*: the spectra in the reconstruction with fully separable sparsity prior; in *blue*: reconstruction with semi-separable sparsity prior.

by the alternate direction method of multipliers (ADMM) is suitable to solve the optimization problem in a short amount of time. In addition to being able to deal with non-differentiable criteria, the ADMM method leads to splitting the full problem in sub-problems that are easier to solve and that may be independent. This straightforwardly gives the opportunity of speeding up the code, *e.g.* by means of parallelization. This possibility remains if other priors are used, *e.g.* to account for a smooth spatial distribution.

To simplify the problem at hand, we considered that complex visibilities have been measured. At optical wavelengths, this is only possible with phase referencing. In order to process most existing interferometric data, we will have to modify the likelihood term $f_{\mathsf{data}}(\boldsymbol{x})$ and use a non-linear method (*i.e.* not the linear conjugate gradients) to update the auxiliary variables $\boldsymbol{z}$. The new algorithm that we proposed, because it treats separately the two cost functions, $f_{\mathsf{data}}(\boldsymbol{z})$ and $f_{\mathsf{prior}}(\boldsymbol{x})$, may however be an efficient alternative to the variable metric method used in MiRA and which treats directly the sum of the cost functions.

ACKNOWLEDGMENTS

This work is supported by the French ANR (*Agence Nationale de la Recherche*), Éric Thiébaut works for the MiTiV project (*Méthodes Inverses de Traitement en Imagerie du Vivant*, ANR-09-EMER-008) and Ferréol Soulez is funded by the POLCA project (*Percées astrophysiques grâce au traitement de données interférométriques polychromatiques*, ANR-10-BLAN-0511).

Our algorithm has been implemented and tested with Yorick (http://yorick.sourceforge.net/) which is available for free.

APPENDIX A. PROXIMITY OPERATORS FOR SPATIAL SPARSITY AND STRICT POSITIVITY

Updating of the variables in the ADMM method consists in solving problems of the form:

$$\min_{\boldsymbol{x}\in\mathbb{X}} \left\{ f(\boldsymbol{x}) + \frac{\alpha}{2} \left\| \boldsymbol{x} - \widetilde{\boldsymbol{x}} \right\|_2^2 \right\} . \tag{24}$$

Solving problem (24) is very close to the job of the so-called proximity operator of the function $f(\boldsymbol{x})$ which is defined as:[18]

$$\mathrm{prox}_f(\widetilde{\boldsymbol{x}}, \alpha) \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \left\{ f(\boldsymbol{x}) + \frac{\alpha}{2} \left\| \boldsymbol{x} - \widetilde{\boldsymbol{x}} \right\|_2^2 \right\} . \tag{25}$$

Proximity operators for non differentiable cost functions like $f_{\mathsf{sparse}}(\boldsymbol{x})$ or $f_{\mathsf{group}}(\boldsymbol{x})$ have already been derived[18] and we simply modified them to account for the additional constraint that $\boldsymbol{x} \in \mathbb{X}$. Since $\mathbb{X}$ is the subset of non-negative vectors of $\mathbb{R}^N$, we denote by:

$$\mathrm{prox}_f^+(\widetilde{\boldsymbol{x}}, \alpha) \stackrel{\text{def}}{=} \underset{\boldsymbol{x}\in\mathbb{R}^N,\ \boldsymbol{x}\geq 0}{\arg\min} \left\{ f(\boldsymbol{x}) + \frac{\alpha}{2} \left\| \boldsymbol{x} - \widetilde{\boldsymbol{x}} \right\|_2^2 \right\}, \tag{26}$$

the modified proximity operator that we use to update the variables $\boldsymbol{x}$ in our algorithm while accounting for $\boldsymbol{x} \in \mathbb{X}$.

A.1 Separable Sparsity

The proximity operator for the ℓ_1 norm is the *soft thresholding* operator:

$$\mathrm{prox}_{\ell_1}(\widetilde{\boldsymbol{x}}, \alpha)_{n,\ell} = \begin{cases} \widetilde{x}_{n,\ell} - 1/\alpha & \text{if } \widetilde{x}_{n,\ell} > +1/\alpha\,; \\ \widetilde{x}_{n,\ell} + 1/\alpha & \text{if } \widetilde{x}_{n,\ell} < -1/\alpha\,; \\ 0 & \text{else}\,. \end{cases} \tag{27}$$

Imposing the positivity is straightforward and yields:

$$\mathrm{prox}_{\ell_1}^+(\widetilde{x}, \alpha)_{n,\ell} = \begin{cases} \widetilde{x}_{n,\ell} - 1/\alpha & \text{if } \widetilde{x}_{n,\ell} > +1/\alpha\,; \\ 0 & \text{else}\,. \end{cases} \tag{28}$$

This shows that if $1/\alpha \geq \max_{n,\ell} \widetilde{x}_{n,\ell}$, the output of the proximity operator is zero everywhere.

A.2 Spatio-Spectral Regularization

First we note that the function $f_{\mathsf{group}}(\boldsymbol{x})$ given by Eq. (10) is separable with respect to the pixel index n. Thus all computations can be done independently for the spectrum of each pixel. For variables $\boldsymbol{x}$ such that the function $f_{\mathsf{group}}(\boldsymbol{x})$ is differentiable, the partial derivatives of $f_{\mathsf{group}}(\boldsymbol{x})$ are:

$$\frac{\partial f_{\mathsf{group}}(\boldsymbol{x})}{\partial x_{n,\ell}} = \frac{x_{n,\ell}}{\beta_n} \quad \text{with} \quad \beta_n \stackrel{\text{def}}{=} \sqrt{\sum\nolimits_{\ell'} x_{n,\ell'}^2}\,. \tag{29}$$

Assuming for the moment that $\beta_n > 0$, the problem of minimizing $f_{\mathsf{group}}(\boldsymbol{x})$ is convex and the solution is obtained by finding the root of the partial derivatives:

$$\begin{aligned} & \frac{\partial}{\partial x_{n,\ell}} \left\{ f_{\mathsf{group}}(\boldsymbol{x}) + \frac{\alpha}{2} \left\| \boldsymbol{x} - \widetilde{\boldsymbol{x}} \right\|_2^2 \right\} = \frac{x_{n,\ell}}{\beta_n} + \alpha\,(x_{n,\ell} - \widetilde{x}_{n,\ell}) = 0 \\ \Longrightarrow \quad & x_{n,\ell} = \frac{\widetilde{x}_{n,\ell}}{1 + (\alpha\,\beta_n)^{-1}} \end{aligned} \tag{30}$$

Solving for β_n yields:

$$\beta_n = \sqrt{\sum\nolimits_{\ell'} x_{n,\ell'}^2} = \frac{\sqrt{\sum_{\ell'} \widetilde{x}_{n,\ell'}^2}}{1 + (\alpha\,\beta_n)^{-1}} \quad \Longrightarrow \quad \beta_n = \sqrt{\sum\nolimits_{\ell'} \widetilde{x}_{n,\ell'}^2} - \frac{1}{\alpha}\,. \tag{31}$$

The non-differentiable case occurs when the expression above yields a value of β_n which is not strictly positive, that is when $\sqrt{\sum_{\ell'} \widetilde{x}_{n,\ell'}^2} \leq 1/\alpha$, in which case the minimum of the cost function is given by $x_{n,\ell} = 0, \forall \ell$. Finally, the proximity operator of $f_{\mathsf{group}}(\boldsymbol{x})$ is:

$$\mathrm{prox}_{f_{\mathsf{group}}}(\widetilde{\boldsymbol{x}}, \alpha)_{n,\ell} = \begin{cases} \left[1 - \left(\alpha \sqrt{\sum_{\ell'} \widetilde{x}_{n,\ell'}^2} \right)^{-1} \right] \widetilde{x}_{n,\ell} & \text{if } \alpha \sqrt{\sum_{\ell'} \widetilde{x}_{n,\ell'}^2} > 1\,; \\ 0 & \text{else.} \end{cases} \tag{32}$$

When $\alpha \sqrt{\sum_{\ell'} \widetilde{x}^2_{n,\ell'}} > 1$, the factor $1 - \left(\alpha \sqrt{\sum_{\ell'} \widetilde{x}^2_{n,\ell'}}\right)^{-1}$ is strictly positive and smaller than one, hence the name *soft compressor* given to this operator.

Requiring that $x_{n,\ell} > 0$ implies that $\beta_n > 0$, since $\alpha > 0$, the solution given by Eq. (30) can only be strictly positive if $\widetilde{x}_{n,\ell}$ is itself strictly positive; otherwise, $x_{n,\ell}$ must be equal to zero. Thus:

$$x_{n,\ell} = \frac{\max(0, \widetilde{x}_{n,\ell})}{1 + (\alpha\, \beta_n)^{-1}}\,, \tag{33}$$

where $\max(0, \widetilde{\boldsymbol{x}})$ is applied element-wise. The rest of the reasonning is similar than with no positivity constraints. The proximity operator of $f_{\mathsf{group}}(\boldsymbol{x})$ modified to account for positivity is finally:

$$\mathrm{prox}^{+}_{f_{\mathsf{group}}}(\widetilde{\boldsymbol{x}}, \alpha) = \mathrm{prox}_{f_{\mathsf{group}}}\left(\max(0, \widetilde{\boldsymbol{x}}), \alpha\right)\,. \tag{34}$$

REFERENCES

[1] Thiébaut, É., "MiRA: an effective imaging algorithm for optical interferometry," in [*Astronomical Telescopes and Instrumentation*], Markus Schöller, William C. Danchi, F. D., ed., **7013**, 70131I–1–70131I–12, SPIE (June 2008).

[2] Buscher, D. F., "Direct maximum-entropy image reconstruction from the bispectrum," in [*IAU Symp. 158: Very High Angular Resolution Imaging*], Robertson, J. G. and Tango, W. J., eds., 91–+ (1994).

[3] Baron, F. and Young, J. S., "Image reconstruction at Cambridge University," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference* **7013**, 144 (July 2008).

[4] Meimon, S., Mugnier, L. M., and le Besnerais, G., "Reconstruction method for weak-phase optical interferometry," *Optics Letters* **30**(14), 1809–1811 (2005).

[5] le Bouquin, J.-B., Lacour, S., Renard, S., Thiébaut, É., and Merand, A., "Pre-maximum spectro-imaging of the Mira star T Lep with AMBER/VLTI," *Astron. Astrophys.* **496**, L1–L4 (2009).

[6] Soulez, F., Thiébaut, É., Bongard, S., and Bacon, R., "Restoration of hyperspectral astronomical data from integral field spectrograph," in [*3rd IEEE WHISPERS*], (2011).

[7] Bongard, S., Soulez, F., Thiébaut, É., and Pécontal, E., "3-D deconvolution of hyper-spectral astronomical data," *Monthly Notices of the Royal Astronomical Society* **418**(1), 258–270 (2011).

[8] Högbom, J. A., "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines," *Astron. Astrophys. Suppl.* **15**, 417–426 (June 1974).

[9] Schwarz, U. J., "Mathematical-statistical Description of the Iterative Beam Removing Technique (Method CLEAN)," *Astron. Astrophys.* **65**, 345–356 (Apr. 1978).

[10] Hofmann, K.-H. and Weigelt, G., "Iterative image reconstruction from the bispectrum," *Astron. Astrophys.* **278**, 328–339 (Oct. 1993).

[11] Donoho, D., "For most large underdetermined systems of linear equations, the minimal ell-1 norm near-solution approximates the sparsest near-solution," *Communications on Pure and Applied Mathematics* **59**(7), 907–934 (2006).

[12] Fornasier, M. and Rauhut, H., "Recovery algorithms for vector valued data with joint sparsity constraints," *SIAM Journal on Numerical Analysis* **46**(2), 577–613 (2008).

[13] Marsh, K. A. and Richardson, J. M., "The objective function implicit in the CLEAN algorithm," *Astron. Astrophys.* **182**, 174–178 (Aug. 1987).

[14] Thiébaut, É., "Image reconstruction with optical interferometers," *New Astronomy Reviews* **53**, 312–328 (2009).

[15] Fessler, J. A. and Sutton, B. P., "Nonuniform Fast Fourier Transforms Using Min-Max Interpolation," *IEEE Trans. Signal Process.* **51**, 560–574 (2003).

[16] Thiébaut, E. and Mugnier, L., "Maximum a posteriori planet detection and characterization with a nulling interferometer," in [*IAU Colloq. 200: Direct Imaging of Exoplanets: Science & Techniques*], Aime, C. and Vakili, F., eds., 547–552 (2006).

[17] Thiébaut, É., "Optimization issues in blind deconvolution algorithms," in [*Astronomical Data Analysis II*], Starck, J.-L. and Murtagh, F. D., eds., **4847**, 174–183, SPIE (2002).

[18] Combettes, P. L. and Pesquet, J.-C., [*Proximal splitting methods in signal processing*], ch. Fixed-Point Algorithms for Inverse Problems in Science and Engineering, 185–212, Springer, New York (2011).

[19] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning* **3**, 1–122 (2010).

[20] Nocedal, J. and Wright, S. J., [*Numerical Optimization*], Springer Verlag, 2nd ed. (2006).

[21] Hestenes, M. R. and Stiefel, E., "Methods of Conjugate Gradients for Solving Linear Systems," *Journal of Research of the National Bureau of Standards* **49**(6), 409–436 (1952).

[22] Jacoby, G. H., Hunter, D. A., and Christian, C. A., "A library of stellar spectra," *Astrophys. J. Suppl.* **56**, 257–281 (1984).

[23] Frigo, M. and Johnson, S. G., "The Design and Implementation of FFTW3," *Proc. IEEE* **93**(2), 216–231 (2005). special issue on "Program Generation, Optimization, and Platform Adaptation".